Quality-Driven Synthesis of Embedded Multi-Mode Control Systems

Samii, Soheil; Eles, Petru; Peng, Zebo; Cervin, Anton

2009

Link to publication

Citation for published version (APA):
Quality-Driven Synthesis of Embedded Multi-Mode Control Systems

Soheil Samii  
Dept. of Computer and Information Science  
Linköping University  
Linköping, Sweden  
sohsa@ida.liu.se

Petru Eles  
Dept. of Computer and Information Science  
Linköping University  
Linköping, Sweden  
petel@ida.liu.se

Zebo Peng  
Dept. of Computer and Information Science  
Linköping University  
Linköping, Sweden  
zpe@ida.liu.se

Anton Cervin  
Dept. of Automatic Control  
Lund University  
Lund, Sweden  
anton@control.lth.se

ABSTRACT
At runtime, an embedded control system can switch between alternative functional modes. In each mode, the system operates by using a schedule and controllers that exploit the available computation and communication resources to optimize the control performance in the running mode. The number of modes is usually exponential in the number of control loops, which means that all controllers and schedules cannot be produced in affordable design-time and stored in memory. This paper addresses synthesis of multi-mode embedded control systems. Our contribution is a method that trades control quality with optimization time, and that efficiently selects the schedules and controllers to be synthesized and stored in memory.

Categories and Subject Descriptors
C.3 [Special-Purpose and Application-Based Systems]: process control systems, real-time and embedded systems; D.4.1 [Operating Systems]: Process Management—scheduling; J.6 [Computer-Aided Engineering]: computer-aided design; J.7 [Computers in Other Systems]: industrial control

General Terms
Algorithms, Design, Performance, Theory

Keywords
control performance, embedded control, multi-mode systems

1. INTRODUCTION AND RELATED WORK
In systems that control several physical plants, mode changes occur at runtime either as responses to external events or at predetermined moments in time. In an operation mode, the system controls a subset of the plants by executing several control applications concurrently—one for each controlled plant in the mode. The application tasks execute periodically (reading sensors, computing control signals, and writing to actuators) on a platform comprising several computation nodes connected to a bus. Such systems have complex timing behavior that leads to bad control performance if not taken into account during controller design [10]. To achieve good control performance in a certain mode, system scheduling must be integrated with controller design (periods and control laws). A mode change means that some of the running control loops are deactivated, some are activated, or both. This leads to a change in the execution and communication demand and, consequently, a new schedule and new controllers must be used to achieve best possible control performance by an efficient use of the resources.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2009, July 26–31, 2009, San Francisco, California, USA.
Copyright 2009 ACM 978-1-60558-497-3/09/07 ...$5.00.

Control-scheduling co-design methods in the literature focus on single-mode control systems. Sampling-period optimization for mono-processor systems with multiple control loops and priority-based scheduling was first studied by Seto et al. [9]. Bini and Cervin improved their work by considering controller delays in the optimization process [2]. Rehbinder and Sandridson proposed optimal control strategies for static scheduling of controllers on a mono-processor system [7]. In our previous work, we proposed a framework [8] for the scheduling and synthesis of controllers on distributed execution platforms.

The contribution of this paper is a scheduling and synthesis method for embedded multi-mode control systems. The number of modes to be considered in the synthesis is usually exponential in the number of control loops and leads to two problems: (1) all modes cannot be synthesized in affordable time, and (2) all synthesized controllers and schedules cannot be stored in memory. With the objective of optimizing the control performance in a multi-mode control system, we address these problems by a limited exploration of the set of modes, followed by a selection of the produced schedules and controllers to store in memory.

2. MULTI-MODE SYSTEMS
Given is a set of plants $P$, indexed by $I_P$, where each plant $P_i$ $(i \in I_P)$ is described by a continuous-time linear model [11]

$$\dot{x}_i = A_i x_i + B_i u_i + v_i, \quad y_i = C_i x_i + e_i, \quad (1)$$

The vectors $x_i$ and $u_i$ are the plant state and controlled input, respectively, and the vector $v_i$ models plant disturbance as a continuous-time white-noise process with intensity $R_{i1}$. The continuous-time output $y_i$ is measured and sampled periodically and is used to produce the control signal $u_i$. The measurement noise $e_i$ in the output is modeled as a discrete-time white-noise process with variance $R_{i2}$. The control signal is updated at discrete time instants and is held constant between two updates (zero-order hold [11]). As an example, let us consider three inverted pendulums $P = \{P_1, P_2, P_3\}$. Each pendulum $P_i$ $(i \in I_P = \{1, 2, 3\})$ is modeled according to Equation 1, with $A_i = \begin{bmatrix} 0 & 1 \\ g/l_i & 0 \end{bmatrix}$, $B_i = \begin{bmatrix} 0 \\ g/l_i \end{bmatrix}^T$, and $C_i = \begin{bmatrix} 1 & 0 \end{bmatrix}$, where $g \approx 9.81 \text{ m/s}^2$ and $l_i$ are the gravitational constant and length of pendulum $P_i$, respectively (all pendulums have equal length, $l_i = 0.2 \text{ m}$). For the plant disturbance and measurement noise, we have $R_{i1} = B_i B_i^T$ and $R_{i2} = 0.1$.

The execution platform comprises a set of nodes $N$, indexed by $I_N$, which are connected by a communication controller to a bus; common communication protocols in automotive systems are TTP [5], FlexRay [4], and CAN [3], which all are supported by our framework. On the execution platform runs a set of applications $A$, indexed by the set $I_A = I_P$. Application $A_i \in A$ $(i \in I_A)$ controls plant $P_i$ and is modeled as a directed acyclic graph $A_i = (T_i, \Gamma_i)$, where the nodes $T_i$, indexed by $I_i$, represent computation tasks and the edges $\Gamma_i \subseteq T_i \times T_i$ represent data dependencies between tasks. Figure 1 shows two nodes $N = \{N_1, N_2\}$ connected to a bus. Three applications $A = \{A_1, A_2, A_3\}$ are mapped to the nodes and control
the three pendulums \( P_1, P_2, \) and \( P_3 \) in the example before. For this example, the task set of application \( A_i \) is \( T_i = \{ \tau_{ic}, \tau_{ic}, \tau_{ia} \} \) for \( i \in \{ s, c, a \} \). The data dependencies are given by the edges \( E = \{ (\tau_{ic}, \tau_{ic}), (\gamma_{ica}, \tau_{ia}) \} \). The sensor task is \( \tau_{ic} \), the controller task is \( \tau_{ic} \), and the actuator task is \( \tau_{ia} \).

The system can run in different operation modes, where each mode is characterized by a set of running applications. A mode is a subset \( M \subseteq \Lambda \), indexed by \( I_M \subseteq I_\Lambda = I_p \), that contains the applications that are running in that mode. The complete set of modes is \( M = 2^\Lambda \) and its cardinality is \( |M| = 2^{|\Lambda|} \). The set of modes \( M \) is a partially ordered set under the subset relation \( \subseteq \). For our example with the three applications in Figure 1, we show in Figure 2 the Hasse diagram of the partially ordered set of modes \( M = 2^\Lambda \). A mode \( M' \in M \) is called a submode of \( M \) if \( M' \subseteq M \). Mode \( M \) is called a supermode of mode \( M' \). We define the set of submodes of \( M \in M \) as \( I_M(M) = \{ M' \in M : M' \subseteq M \} \). Similarly, the set of supermodes of \( M \) is denoted \( \overline{I}_M(M) \).

The idle mode is the empty set \( \emptyset \), which indicates that the system is inactive, whereas the mode \( \Lambda \) indicates that all applications are running. It can be the case that certain modes do not occur at runtime—for example, because certain plants are never controlled concurrently. Let us therefore introduce the set of functional modes \( \mathcal{M}^{\text{func}} \subseteq M \) that can occur during execution. Modes \( \mathcal{M}^{\text{vist}} = M \setminus \mathcal{M}^{\text{func}} \) do not occur at runtime and are therefore called virtual modes. In a given mode \( M \in \mathcal{M} \), an application \( A_i \in M \) (\( i \in I_M \)) releases jobs for execution periodically with the period \( h^M \). Job \( q \) of task \( \tau_{ij} \) is denoted \( \tau(q)_{ij} \). For a message \( \gamma_{ijk} = (\tau_{ij}, \tau_{ik}) \in \Gamma_i \), the message instance produced by job \( \tau_{ij}^{(q)} \) is denoted \( \gamma(q)_{ij} \). An edge \( \gamma_{ijk} = (\tau_{ij}, \tau_{ik}) \in \Gamma_i \) means that the earliest start time of job \( \tau_{ik}^{(q)} \) is when \( \tau(q)_{ij} \) has completed its execution and the produced data (i.e., \( \gamma(q)_{ij} \)) has been communicated to \( \tau_{ik}^{(q)} \). We define the hyperperiod \( h_M \) of \( M \) as the least common multiple of the periods \( h^M \) for \( i \in I_M \).

Each task is mapped to a node; in Figure 1, task \( \tau_{ia} \) is mapped to \( N_1 \) and \( \tau_{ic} \) is mapped to \( N_2 \). A message between tasks that are mapped to different nodes is sent on the bus. For a message \( \gamma_{ijk} \) between two nodes, we denote with \( c_{ij} \) the communication time when there are no conflicts on the bus. We model the execution time of task \( \tau_{ij} \) as a stochastic variable \( c_{ij} \) with probability function \( \xi_{ij} \). We assume that the execution time of \( \tau_{ij} \) is bounded by given best-case and worst-case execution times. In Figure 1, the execution times (constant in this example) and communication times for the tasks and messages are given in milliseconds in parentheses. We support static-cyclic and priority-based scheduling of tasks and messages [5]. In static-cyclic scheduling, the period of the schedule is the hyperperiod \( h_M \) of the applications in the running mode \( M \). The schedule determines the start times of the jobs and message instances that are released within a hyperperiod of the mode. Further, the schedule must satisfy precedence constraints, which are given by the data dependencies, and account for the worst-case execution times of the tasks and the communication times of the messages. For preemptive priority-based scheduling, the tasks and messages are scheduled at runtime based on fixed priorities that are decided at design time.

As an example of a static-cyclic schedule, let us consider the system in Figure 1 and the schedule in Figure 3. All times are given in milliseconds in the discussion that follows. The schedule is constructed for mode \( M = \Lambda \) (i.e., for the mode in which all three applications are running concurrently) and for the periods \( h_1^M = 40, h_2^M = 20, \) and \( h_3^M = 40 \). The period of the schedule is \( h_M = 40 \) (the hyperperiod of \( M \)). Considering the periods of the applications, the jobs to be scheduled on node \( N_1 \) are \( \tau_{ia}^{(1)}, \tau_{ic}^{(1)}, \tau_{ic}^{(2)}, \tau_{ic}^{(3)}, \tau_{ic}^{(4)}, \tau_{ic}^{(5)}, \) and \( \tau_{ic}^{(6)} \). The jobs on node \( N_2 \) are \( \tau_{ia}^{(1)}, \tau_{ic}^{(1)}, \tau_{ic}^{(2)}, \tau_{ic}^{(3)}, \tau_{ic}^{(4)}, \) and \( \tau_{ic}^{(5)} \). The message transmissions on the bus are \( \gamma_{ia}^{(1)} \), \( \gamma_{ia}^{(1)} \), \( \gamma_{ia}^{(2)} \), \( \gamma_{ia}^{(3)} \), and \( \gamma_{ia}^{(4)} \). The schedule in Figure 3 is shown with three rows for node \( N_1 \), the bus, and node \( N_2 \), respectively. The small boxes depict task executions and message transmissions. The white, grey, and black boxes show the execution of applications \( A_1, A_2, \) and \( A_3 \), respectively. Each box is labeled with an index that indicates a task or message, and with a number that specifies the job or message instance. The black box labeled \( c(1) \) shows that the execution of job \( \tau_{ic}^{(1)} \) on node \( N_2 \) starts at time 3 and completes at time 20. The grey box labeled \( sc(2) \) shows the second message between the sensor and controller task of \( A_2 \).

An implementation of a mode \( M \in \mathcal{M} \) comprises the period \( h_M^M \) and control law \( \gamma_M \) of each control application \( A_i \in M \) (\( i \in I_M \)), and the schedule (or priorities) for the tasks and messages in the mode. We denote with \( \text{mem}_M \) the memory consumption on node \( N_d \in N \) (\( d \in I_N \)) of the implementation of \( M \). An implementation of a mode can serve as an implementation of all its submodes. This means that the system can run in a submode \( M' \in \mathcal{M} (M' \neq \emptyset) \) with the same controllers (periods and control laws) and schedule as for mode \( M \)—for example, by not running the applications in \( \mathcal{M} \setminus M' \) or by not writing to their outputs. However, to achieve better performance in mode \( M' \), a customized set of controllers and schedule for submode \( M' \) can exploit the available computation and communication resources that are now not used by the other applications \( \mathcal{M} \setminus M' \). To have a correct implementation of the whole system, for each functional mode \( M \in \mathcal{M}^{\text{func}} \), there must exist an implementation in memory, or there must exist an implementation of at least one of its supermodes. The set of implemented modes is \( \mathcal{M}^{\text{impl}} \subseteq \mathcal{M} \setminus \{ \emptyset \} \).

### 3. CONTROL QUALITY AND SYNTHESIS

The quality of a controller \( A_i \) for plant \( P_i \) (Equation 1) is given by the quadratic cost [11]

\[
J_i = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left\{ \int_0^T \left[ x_i - u_i \right]^T Q_i \left[ x_i - u_i \right] \, dt \right\}
\]  

The matrix \( Q_i \), which is given by the designer, is a positive semi-definite matrix with weights for the magnitude of the plant states and the control signals \( E \{ \cdot \} \) denotes the expected value of a stochastic variable. For a given sampling period \( h_i \) and a given constant sensor–actuator delay (i.e., the time elapsed between sampling the output \( y_i \) and updating the controlled input \( u_i \)), it is possible to find the control law \( u_i \) that minimizes \( J_i \). [11]. The cost \( J_i \) of a controller is increased (its quality is
degraded) if the sensor–actuator delay at runtime is different from what was assumed during control-law synthesis, or if this delay is not constant (i.e., there is jitter). Given the delay characteristics, we compute this cost $J_i$ with Jitterbug [6].

We have developed a framework [8] to synthesize controllers and schedule their execution and communication for single-mode systems. In this paper, we use the framework to obtain an implementation of a certain mode $M \in M \setminus \{\emptyset\}$ such that the total control cost of the mode $J^M = \sum_{i \in I_M} J_i^M$ is minimized, where $J_i^M$ is the cost of controller $A_i$. The produced mode implementation comprises the periods $h_i^M$, control laws $u_i^M$, and system schedule (schedule table or priorities). We also obtain the memory consumption $\text{mem}^M \in N$ of the implementation on each computation node $N_i \in N$ ($d \in I_M$).

4. MOTIVATIONAL EXAMPLE

In this section, we shall highlight the two problems that are addressed in this paper: (1) the time complexity of the synthesis of embedded multi-mode control systems, and (2) the memory complexity of the storage of produced mode implementations. Let us consider our example in Section 2 with three applications $\Lambda = \{A_1, A_2, A_3\}$ that control the three inverted pendulums $P = \{P_1, P_2, P_3\}$. For the computation of the control cost in Equation 2, we have used $Q_i = \text{diag}(C_i^T C_i, 0.002)$ as the weight matrix for each inverted pendulum $P_i$.

By using our framework for controller scheduling and synthesis [8], we synthesized an implementation of mode $M_{123} = \Lambda$ with the periods $h_{123}^M = 40$, $h_{13}^M = 20$, and $h_{23}^M = 40$ and the schedule in Figure 3. The outputs of the plants are sampled periodically without jitter (e.g., by some mechanism that stores the sampled data in buffers). The actuations of $A_1$ and $A_3$ finish at times 40 and 28, respectively. Because the schedule is periodic, the delay from sampling to actuation is 40 in each instance of $A_1$ (i.e., constant). Similarly, application $A_3$ has the constant sampling–actuation delay 28. Two instances are scheduled for application $A_2$. The first actuation finishes at time 20, whereas the second actuation finishes at time 38. By considering the sampling period 20, we note that the sampling–actuation delay of $A_2$ during periodic execution of the schedule is either 20 or 18. With this implementation, we obtained the individual control costs $J_{123}^M = 2.98$, $J_{13}^M = 1.19$, and $J_{23}^M = 2.24$. Table 1 contains the individual controller costs of the modes considered in this section.

Let us now consider mode $M_{12} = \{A_1, A_2\}$ in which $A_3$ is not executing. The system can operate in the new mode by using the schedule and control laws from mode $M_{123}$. This can be done by not writing to the actuators of $A_3$ or by omitting the execution and communication of $A_3$. By using the implementation of mode $M_{123}$, the overall control cost of mode $M_{12}$ is $J_{12}^M = J_{123}^M + J_{2}^M = 2.98 + 1.19 = 4.17$. This cost can be reduced because, compared to mode $M_{123}$, there is now more computation and communication power available for applications $A_1$ and $A_2$. Thus, it is worth investigating whether a better implementation (e.g., with reduced periods and delays) can be produced for mode $M_{12}$. By running the synthesis of $M_{123}$, we obtained an implementation with the periods $h_{12}^M = h_{2}^M = 30$ and the schedule in Figure 4. Note from the new schedule that both the period and delay of $A_1$ have been reduced. The costs of $A_1$ and $A_2$ with the new implementation are $J_{12}^M = 1.60$ and $J_{2}^M = 1.42$ (Table 1). The cost of $A_1$ is reduced significantly as a result of the reduction in sampling period and delay. The sampling period of $A_2$ is increased, which results in a small increase in the cost of $A_2$. This cost increase is accepted because it makes possible a significant quality improvement of $A_1$, which leads to a significant cost reduction of mode $M_{12}$ to 3.02.

To achieve the best control performance, implementations of all functional modes have to be produced at design time. However, the number of modes is exponential in the number of control loops that run on the platform. Thus, even if some modes are functionally excluded, implementations of all possible functional modes cannot be produced in affordable time (except cases with small number of control loops). Let us consider that we can only run the synthesis of three modes at design time. Modes $M_{123}$ and $M_{12}$ have already been discussed. Considering the third mode to be $M_{13}$, the synthesis result in the costs $J_{13}^M = 1.60$ and $J_{23}^M = 1.95$ (Table 1). When using the implementation of $M_{123}$, the costs of $A_1$ and $A_2$ are 2.98 and 2.24, respectively. By running $M_{13}$ with the new implementation thus leads to a significant improvement in control performance, compared to when using the implementation of $M_{123}$. At runtime, the system has implementations of the modes $M_{13}^{\text{impl}} = \{M_{13}, M_{123}, M_{12}\}$ in memory. For modes that are not implemented, the system chooses an implementation at runtime based on the three available implementations. For example, mode $M_{2} = \{A_2\}$ does not have an implementation but can run with the implementation of either $M_{12}$ or $M_{123}$. The cost of $A_2$ when running with the implementation of $M_{12}$ is $J_{2}^{M_{12}} = 1.42$, whereas it is $J_{2}^{M_{123}} = 1.19$ for the implementation of $M_{123}$. Thus, at runtime, the implementation of $M_{123}$ is chosen to operate the system in mode $M_{2}$.

We now consider that memory limitations in the platform imply that we can only store implementations of two modes out of the three modes in $M^{\text{impl}}$. We cannot use the implementation of $M_{12}$ or $M_{13}$ to run the system in mode $M_{123}$ and thus we cannot remove its implementation. As we discussed in the beginning of this section, the total control cost when running mode $M_{12}$ with the implementation of $M_{123}$ is 4.17, compared to the total cost 3.02 when running with the implementation of $M_{12}$. The implementation of $M_{12}$ thus gives a total cost reduction of 1.15. If, on the other hand, $M_{13}$ runs with the implementation of $M_{123}$, the total cost is $J_{13}^M + J_{23}^M = 2.98 + 2.24 = 5.22$. If $M_{13}$ runs with its produced implementation, the total cost is $J_{13}^M + J_{23}^M = 1.60 + 1.95 = 3.55$. This gives a total cost reduction of 1.67, which is better than the reduction obtained by the implementation of $M_{12}$. Thus, in the presence of memory limitations, the implementations of $M_{123}$ and $M_{13}$ should be stored in memory to achieve the best control performance. In this discussion, we have assumed that $M_{12}$ and $M_{13}$ are equally important. However, if $M_{12}$ occurs more frequently than $M_{13}$, the cost improvement of the implementation of $M_{12}$ becomes more significant. In this case, the best selection could be to exclude the implementation of $M_{13}$ and store implementations of $M_{123}$ and $M_{12}$ in memory. As the last example, let us consider that $M_{123}$ is a virtual mode (i.e., it does not occur at runtime). Let $M^{\text{impl}} = \{M_{12}, M_{13}, M_{23}\}$ be the set of modes with produced implementations. We have run the synthesis of mode $M_{23} = \{A_2, A_3\}$ and obtained a total cost of 3.31. Let us assume that the three produced implementations of the modes in $M^{\text{impl}}$ cannot be all stored in memory. If the implementation of $M_{23}$ is removed,
for example, there is no valid implementation of the functional mode \( M_{123} \) in memory. To solve this problem, we implement the virtual mode \( M_{123} \). Its implementation can be used to run the system in all functional modes—however, with degraded control performance. The available memory allows us to further store the implementation of one of the modes \( M_{122}, M_{112}, \) or \( M_{23} \). We choose the mode implementation that gives the largest cost reduction, compared to the implementation with \( M_{123} \). By looking in Table 1 and with the cost reductions in our discussions earlier in this example, we conclude that \( M_{13} \) gives the largest cost reduction among the modes in \( M^{\text{impl}} \). The best control performance under these tight memory constraints is achieved if the virtual mode \( M_{123} \) and functional mode \( M_{13} \) are implemented and stored in memory. This shows that memory limitations can lead to situations in which virtual modes must be implemented to cover a set of functional modes.

5. PROBLEM FORMULATION

In addition to the plants, control applications, and their mapping to an execution platform (Section 2), the inputs to the problem that we address are

- a set of functional modes\(^1\) \( M^{\text{func}} \subseteq M \) that can occur at runtime (the set of functional modes is \( M^{\text{virt}} = M \setminus M^{\text{func}} \)),
- a weight\(^2\) \( w_M > 0 \) for each mode \( M \in M^{\text{func}} \) (\( w_M = 0 \) for \( M \in M^{\text{virt}} \)), and
- the available memory in the platform, modeled as a memory limit \( \text{mem}^{\text{max}} \in \mathbb{N} \) for each node \( d \in \mathcal{D} \).

The outputs are implementations of a set of modes \( M^{\text{impl}} \subseteq M \setminus \{ \emptyset \} \). For each functional mode \( M \in M^{\text{func}} \setminus \{ \emptyset \} \), there must exist an implementation (i.e., \( M \in M^{\text{impl}} \)) or a supermode implementation (i.e., \( M^{\text{impl}} \cap \overline{M} \)). If \( M \in M^{\text{impl}} \), the cost \( J^{M} \) is given from the scheduling and synthesis step that produces the implementation (Section 3). If \( M \notin M^{\text{impl}} \) and \( M \neq \emptyset \), the cost \( J^{M} \) is given by the available supermode implementations: Given an implemented supermode \( M' \in M^{\text{impl}} \cap \overline{M} \), the cost of \( M \) when using the implementation of \( M' \) is

\[
J^{M}(M') = \sum_{i \in M' \cap \mathcal{M}(M)} J^{M}(i) = \sum_{i \in M' \cap \mathcal{M}(M)} J^{M}(i) - \sum_{i \in M' \cap \mathcal{M}(M)} J^{M}(i).
\]

At runtime, we use the supermode implementation that gives the smallest cost, and thus

\[
J^{M} = \min_{M' \in M^{\text{impl}} \cap \overline{M}(M)} J^{M}(M').
\]

The cost of the idle mode \( \emptyset \) is defined as \( J^{\emptyset} = 0 \). The objective is to find a set of modes \( M^{\text{impl}} \) and synthesize them such that their implementations can be stored in the available memory of the platform. The cost to be minimized is

\[
J_{\text{tot}} = \sum_{M \in M^{\text{func}}} w_M J^{M}.
\]

6. SYNTHESIS APPROACH

Our synthesis approach consists of two sequential parts. First, we synthesize implementations for a limited set of functional modes (Section 6.1). Second, we select the implementations to store under given memory constraints, and if needed, we produce virtual-mode implementations (Section 6.2).

---

\(^1\)Due to the large number of modes, it can be nonpractical (or even impossible) to explicitly mark the set of functional modes. In such cases, however, the designer can indicate control loops that cannot be active in parallel due to functionality restrictions. Then, the set of functional modes \( M^{\text{func}} \) includes all modes except those containing two or more mutually exclusive controllers. If no specification of functional modes is made, it is assumed that \( M^{\text{func}} = M \).

\(^2\)It can be impossible for the designer to assign weights to all functional modes equally. An alternative is to assign weights to some particularly important and frequent modes (all other functional modes get a default weight). Another alternative is to correlate the weights to the occurrence frequency of modes (e.g., obtained by simulation).
1. If \( \sum_{M \in M^{\text{impl}}_b} \text{mem}_M^b > \text{mem}_{\text{max}} \) for some \( d \in \mathbb{N} \):
   
   (a) Find \( M' \subseteq \bigcup_{M \in M^{\text{impl}}_b} M^+(M) \) with smallest size such that, for each \( M \in M^{\text{impl}}_b \), \( M' \cap M^+(M) \neq \emptyset \).
   
   (b) For each \( M' \in M' \), synthesize \( M' \) and set \( M'^{\text{impl}} := M^{\text{impl}} \cup M'(M') \).
   
   (c) Repeat Step 1

2. Find \( b_M \in \{0, 1\} \) for each \( M \in M^{\text{impl}}_b \) such that the cost in Equation 5 is minimized and the constraint in Equation 6 is satisfied for each \( d \in \mathbb{N} \) and \( b_d = 1 \) for \( M \in M^{\text{impl}}_b \).

**Figure 6: Mode-selection approach**

This approach is used to select the mode implementations to store in the available memory such that the control quality is optimized. This selection is done by solving a linear program with binary variables: Given the set of implemented modes, possibly including virtual modes from Step 1 in Figure 6, let us introduce a binary variable \( b_M \in \{0, 1\} \) for each mode \( M \in M^{\text{impl}}_b \). If \( b_M = 1 \), it means that the implementation of \( M \) is selected to be stored in memory, whereas if it is not stored if \( b_M = 0 \). For a correct implementation of the multi-mode system, we must store the implementation of each top mode \( M \in M^{\text{impl}}_b \), and thus \( b_M = 1 \) for those modes. To select the other mode implementations to store, we consider the obtained cost reduction by implementing a mode. Given \( M \in M^{\text{impl}}_b \), this cost reduction is

\[
\Delta J^M = \min_{M' \subseteq M^{\text{impl}}_b \setminus M^{\text{impl}}_b} \Delta J^M(M'),
\]

where \( \Delta J^M(M') \) is given in Step ii in Figure 5. By removing a mode \( M \in M^{\text{impl}}_b \setminus M^{\text{impl}}_b \) (i.e., setting \( b_M = 0 \)), we lose this cost improvement. The cost to be minimized is the overall loss of cost reductions for mode implementations that are removed from \( M^{\text{impl}}_b \setminus M^{\text{impl}}_b \). This cost is formulated as

\[
\sum_{M \in M^{\text{impl}}_b \setminus M^{\text{impl}}_b} (1 - b_M) \Delta J^M(M).
\]

The memory consumption of an implementation of mode \( M \) is \( \text{mem}^b_M \) on node \( N_d \in \mathbb{N} \). This information is given as an output of the controller scheduling and synthesis (Section 3). The memory constraint on node \( N_d \in \mathbb{N} \) is

\[
\sum_{M \in M^{\text{impl}}_b} b_M \text{mem}_M^b \leq \text{mem}_{\text{max}}.
\]

6.2 Control-Quality versus Memory Consumption

Given are implementations of functional modes \( M^{\text{impl}}_b \) as a result of the first part. We shall now discuss how to select the implementations to store such that given memory constraints are satisfied and the system can operate in any functional mode with the stored implementations. Note that the set of top functional modes \( M^{\text{func}}_b \) must have implementations in memory. These implementations can be used to store the system in any of the other functional modes. If the implementations of the top functional modes can be stored in the available memory, we do not need to implement virtual modes. The remaining problem is to additionally select the remaining implementations to store in memory. If, however, the implementations of the top functional modes cannot be stored, we must produce implementations of virtual modes. These implementations should cover a large amount of functional modes. An implementation of a virtual supermode can replace several implementations and, therefore, save memory space but with degraded control performance in the modes with replaced implementations.

To select the mode implementations to store, we propose the two-step approach outlined in Figure 6. We first check whether the implementations of the functional top modes \( M^{\text{impl}}_b = M^{\text{func}}_b \) can be stored in the available memory on each computation node \( N_d \in \mathbb{N} \). If not, we proceed by selecting virtual modes to implement. In Step (a), we find among the immediate submodes of the top modes \( M^{\text{impl}}_b \) (the submodes are virtual modes) a subset \( M' \) with minimal size that contains a supermode for each top mode. In Step (b), we produce implementations of the chosen virtual modes \( M' \) and add them to the set of implemented modes \( M^{\text{impl}}_b \). The top modes of \( M^{\text{impl}}_b \) is now \( M' \). We go back to Step 1 and check whether the implementations of the new top modes can be stored in memory. If not, we repeat Steps (a) and (b) and consider larger virtual modes until we find a set of virtual modes that cover all functional modes and that can be implemented in the available memory. After this step, we know that the implementations of \( M^{\text{impl}}_b \) can be stored in the available memory and that they are sufficient to operate the system in any functional mode.

The second step selects mode implementations to store in the available memory such that the control quality is optimized. This selection is done by solving a linear program with binary variables: Given the set of implemented modes, possibly including virtual modes from Step 1 in Figure 6, let us introduce a binary variable \( b_M \in \{0, 1\} \) for each mode \( M \in M^{\text{impl}}_b \). If \( b_M = 1 \), it means that the implementation of \( M \) is selected to be stored in memory, whereas if it is not stored if \( b_M = 0 \). For a correct implementation of the multi-mode system, we must store the implementation of each top mode \( M \in M^{\text{impl}}_b \), and thus \( b_M = 1 \) for those modes. To select the other mode implementations to store, we consider the obtained cost reduction by implementing a mode. Given \( M \in M^{\text{impl}}_b \), this cost reduction is

\[
\Delta J^M = \min_{M' \subseteq M^{\text{impl}}_b \setminus M^{\text{impl}}_b} \Delta J^M(M'),
\]

where \( \Delta J^M(M') \) is given in Step ii in Figure 5. By removing a mode \( M \in M^{\text{impl}}_b \setminus M^{\text{impl}}_b \) (i.e., setting \( b_M = 0 \)), we lose this cost improvement. The cost to be minimized is the overall loss of cost reductions for mode implementations that are removed from \( M^{\text{impl}}_b \setminus M^{\text{impl}}_b \). This cost is formulated as

\[
\sum_{M \in M^{\text{impl}}_b \setminus M^{\text{impl}}_b} (1 - b_M) \Delta J^M(M).
\]

The memory consumption of an implementation of mode \( M \) is \( \text{mem}^b_M \) on node \( N_d \in \mathbb{N} \). This information is given as an output of the controller scheduling and synthesis (Section 3). The memory constraint on node \( N_d \in \mathbb{N} \) is

\[
\sum_{M \in M^{\text{impl}}_b} b_M \text{mem}_M^b \leq \text{mem}_{\text{max}}.
\]

**7. EXPERIMENTAL RESULTS**

We have conducted experiments to evaluate our proposed approach. We created 100 benchmarks using a database of inverted pendulums, ball and beam processes, DC servos, and harmonic oscillators [11]. Such plants are acknowledged as representatives of realistic control problems and are used extensively for experimental evaluation. We considered 10 percent of the modes to be virtual for benchmarks with 6 or more control loops. We generated 3 to 5 tasks for each control application; thus, the number of tasks in our benchmarks varies from 12 to 60. The tasks were mapped randomly to platforms comprising 2 to 10 nodes. Further, we considered uniform distributions of the execution times of tasks. The best-case and worst-case execution times of the tasks were chosen randomly from 2 to 10 milliseconds and the communication times of messages were chosen from 2 to 4 milliseconds.

The first set of experiments evaluate the synthesis heuristic in Section 6.1. The synthesis was run for each benchmark and for different values of the improvement factor \( \lambda \). All experiments were run on a PC with a quad-core CPU at 2.2 GHz, 8 Gb of RAM, and running Linux. For each value of \( \lambda \) and for each benchmark, we computed the obtained cost after the synthesis. This cost is \( J_{\text{tot}} = \sum_{M \in M^{\text{func}}} \text{mem}_M J^M \) (Equation 4) and, because of the large number of modes, we could compute...
it in affordable time for benchmarks with at most 12 control loops. It is very important to note that the time-consuming calculation of the cost is only needed for the experimental evaluation and is not part of the heuristic. We used the values 0, 0.1, 0.3, and 0.5 for the improvement factor λ. Because of long runtimes, we could run the synthesis with λ = 0 (exhaustive synthesis of all functional modes) and λ = 0.1 for benchmarks with at most 7 control loops. As a baseline for the comparison, we used the cost $J^\text{syn}_{\text{tot}}$ obtained when synthesizing only the top functional modes $M^\text{func}_{\text{top}}$. For each benchmark and for each value of λ, we computed the achieved cost improvement as the relative difference \[ \Delta J^\text{(λ)}_{\text{tot}} = \left( J^\text{tot}_{\text{syn}} - J^\text{(λ)}_{\text{tot}} \right) / J^\text{tot}_{\text{syn}}. \] Figure 7 shows the obtained cost improvements. The vertical axis indicates the average value of the relative cost-improvement $\Delta J^\text{(λ)}_{\text{tot}}$ for benchmarks with number of control loops given on the horizontal axis. The results show that the achieved control performance becomes better with smaller values of λ (i.e., a more thorough exploration of the set of modes). We also observe that the improvement is better the larger number of control loops in the multi-mode system. This demonstrates that for larger number of control loops, it is important to additionally synthesize other modes than the top functional modes. The experiments also show that good quality results can be obtained also with larger values of λ, which provide affordable runtimes for even larger examples. The runtimes for the synthesis heuristic are shown in Figure 8. We show the average runtimes in seconds for the different values of the improvement factor λ and for each dimension of the benchmarks (number of control loops). Figures 7 and 8 illustrate how the designer can trade off quality of the synthesis with optimization time, depending on the size of the control system. To further illustrate the scaling of synthesis time, we mention that a system consisting of 20 control loops on 12 nodes has been synthesized with $\lambda = 0.5$ in 38 minutes.

In the second set of experiments, we evaluate the mode-selection approach. We have run the mode-synthesis heuristic (Section 6.1) with $\lambda = 0.3$. Let us denote with $\text{mem}^{\text{req}}$ the amount of memory needed to store the generated mode implementations on each node $N_i$. We have run the mode selection (Section 6.2) considering a memory limitation of $0.7\text{mem}^{\text{req}}$ for each node. To solve the linear program given by Equations 5 and 6, we used the eplex library for mixed integer programming in ECLiPSe [1]. The average runtimes are shown in seconds in Figure 9. Note that the mode selection is optimal if all top functional modes can be stored in memory. The selection of modes to store in memory is performed only once as a last step of the synthesis, and as the figure shows, the optimal selection is found in neglectable runtime compared to the overall runtime of the multi-mode system synthesis. To study the quality degradation as a result of memory limitations, we used a benchmark with 12 control loops running on 10 nodes. We have run the mode synthesis for three scenarios: First, we considered no memory limitation in the platform, which resulted in a cost improvement of 48.5 percent, relative to the cost obtained with the baseline approach (synthesis of only top functional modes). Second, we considered that only 70 percent of the memory required by the produced implementations can be used. As a result of the mode selection, all top functional modes could be stored, leading to a cost improvement of 41.4 percent (a degradation of 7.1 percent compared to the first scenario without memory constraints). For the third and last scenario, we considered memory limitations such that the implementations of the top functional modes cannot be all stored in memory. After the mode-selection approach, including the synthesis of virtual modes, we obtained a solution with a cost improvement of only 30.1 percent (a degradation of 18.4 percent compared to the case without memory constraints).

8. CONCLUSIONS

We addressed control-performance optimization for embedded multi-mode control systems. The main design difficulty is raised by the potentially very large number of possible modes. In this context, an appropriate selection of the actual modes to be implemented is of critical importance. We presented our synthesis approach that produces schedules and controllers for an efficient deployment of embedded multi-mode control systems.

9. REFERENCES